TI 2005-107/3
Tinbergen Institute Discussion Paper

Spatial Activity and Labour Market Patterns: A Connectivity Analysis of Commuting Flows in Germany

*Giovanni Russo**
*Aura Reggiani***
*Peter Nijkamp****

* IRES FvG, Trieste, Italy;
** University of Bologna, Bologna, Italy;
*** Vrije Universiteit Amsterdam, and Tinbergen Institute.

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

SPATIAL ACTIVITY AND LABOUR MARKET PATTERNS: A CONNECTIVITY ANALYSIS OF COMMUTING FLOWS IN GERMANY

Giovanni Russo*, Aura Reggiani and Peter Nijkamp*****

* IRES FvG, Via della Fonderia 3, 34129 Trieste, Italy; email: giovanni_russo@alice.it

** Department of Economics, Faculty of Statistics, University of Bologna, Piazza Scaravilli, 2, 40126 Bologna, Italy; e-mail: aura.reggiani@unibo.it

*** Department of Spatial Economics, Faculty of Economics, Free University, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands; e-mail: pnijkamp@feweb.vu.nl

ABSTRACT

The spatial activity patterns of firms in a multi-regional system are closely connected with the structure and evolution of regional labour markets. Based on an extensive data set (cross-section) on commuting flows in Germany, this paper aims to identify the relationship between entrepreneurial activity and spatial labour markets, by employing in particular the concept of 'entrepreneurial city'. A network connectivity model is adopted to assess connectivity patterns, using the power-law and exponential law as a statistical test framework, in order to detect the presence of economic activity hubs that may resemble the concept of entrepreneurial cities. Various results are presented and interpreted in the final part of the paper.

PN202GRAR

1. Scoping the Scene

The emergence of regional (or urban) growth and development is critically dependent on prominent explanatory factors related to innovation and entrepreneurship. It is nowadays widely assumed that innovative activity of entrepreneurs forms the key for understanding spatial dynamics. Studies on entrepreneurship have gained much popularity in recent years, as the "entrepreneurial hero" is recognized as the key actor in an industrial system. In an ICT society, the radius of innovative firms may extend the global level, so that spatial (regional, urban) networks become the playground for many modern entrepreneurs. Consequently, the geography of innovation does not only have a local dimension, but, through the existence of networks, also has a wide geographical coverage. This observation is also clearly reflected in the new economic geography as well as in the principles of the learning economy. Spatial dynamics, modern network configurations and innovative entrepreneurial activity are clearly interrelated phenomena that form the foundations of economic growth and development, both nationally and regionally.

Successful entrepreneurship is reflected in above-average economic growth. Strong economic growth is a key ingredient for national and regional welfare. In fact, high economic growth and the ensuing high rates of job creation are instrumental for sustaining low unemployment rates. Economic growth is not, however, emerging in a wonderland of no geographical dimensions, but is clearly mapped out in a space-time framework. The geographical development of business activity and of regional labour markets are two closely intertwined phenomena.

Several empirical illustrations may clarify the above premise. For example, strong economic growth in the US has often been attributed to technological progress in the ICT industries which has lowered the price of ICT capital, thus favouring investment and capital accumulation in ICT-user industries (Jorgenson and Stiroh 2000; Jorgenson et al. 2003).

In the economics literature, technological innovation is often claimed to be a critical success factor for economic growth, but it is by no means a sufficient factor. Besides technological progress in the ICT industries, investments in human and physical capital – supported by pro-competitive macroeconomic policies – have also been proven to be important sources of growth (OECD 2003). In fact, it has been found that the entry of innovative firms is a particularly important source of productivity growth (OECD 2003).

Growth is partly the result of efficiency increases of incumbent firms in a competitive economy, and partly the result of new entrants challenging the established firms. The entry of new firms, known as 'start-ups', is intimately linked to innovative entrepreneurial activity. The link between entry rates and economic growth is usually very robust, and it can be found at various levels of analysis (firm, region, and nation). At the firm level, young firms, small firms and new firms – the result of entrepreneurial activity – positively affect economic growth, since they all have higher growth rates than old firms, large firms and incumbent firms, respectively; however, young firms are less likely to survive than large firms (Audretsch 1995; Caves 1998; Sutton 1997).

Moreover, high entry rates of new firms have a positive impact on economic growth at the regional level, albeit with a considerable time lag (Audretsch and Fritsch 2002; van Stel and Diephuis 2004). Interestingly, among German regions, the link between entrepreneurial activity and regional economic growth emerged only during the 1990s, as a consequence of the surge in international outsourcing of large German manufacturing firms, spurred by globalization (Audretsch and Fritsch 2003).

Finally, entry rates of new and innovative firms – the consequence of entrepreneurial activity – are also found to have a positive impact on aggregate economic growth and a negative impact on unemployment (Audretsch and Thurik 2001; Carree and Thurik 2002; OECD 2003; van Stel and Diephuis 2004). The entrepreneur turns out to be a central actor in economic growth strategies.

Because new entrepreneurial activity (start-ups) and economic growth are firmly interlocked, we observe in the US a renewed interest in the role of the entrepreneur and in policy measures designed to stimulate entrepreneurship (Reynolds 1999). This issue is an important one, because, in fact, the presence of various types of externalities implies that economic systems as such would rarely produce the optimal level of entrepreneurial activity. However, the selection of a correct mix of policies can effectively foster entrepreneurial activity. Policy intervention is fraught with dangers: the selection of a poor policy mix can seriously curb the degree of entrepreneurial activity (Davidsson and Henrekson 2002).

The justification for (careful) policy intervention results from the presence of market failures in four distinct areas: a) network externalities; b) knowledge externalities; c) failure externalities; and d) learning externalities (Audretsch 2004). All four externalities are conducive to spatial agglomeration:

a) the first externality refers to the proximity of complementary firms or individuals;
b) the second externality refers to knowledge spillovers, since firms benefit from the proximity of similar firms from which they can learn;
c) the third externality refers to the economic value created for third parties by new firms going out of business. Ideas and projects pursued by non-successful firms are often subsequently incorporated in the products and projects of successful firms;

d) the fourth externality refers to demonstration and imitation, as potential entrepreneurs observe the activity – and the results from entrepreneurial activity – and this may spur them into action, thus generating new waves of innovating firms.

In other words, the entrepreneur is the key agent who realizes knowledge spillovers from the source to final users (consumers, investment, and export goods), and to others who may use them as intermediate products (Audretsch and Thurik 2001).

The relationship between growth and space deserves some more attention. Space – or distance – can act as an impediment to growth, but may also function as a factor accelerating economic development. Spatial concentration of economic activity has often been observed in the empirical literature as the result of economies of density or agglomeration advantages (Glaeser 1998). In this process, cities act as catalysts for entrepreneurial activity, because entrepreneurial activity does not take place in a vacuum. In fact, besides the proximity of other fellow entrepreneurs, entrepreneurs in the knowledge economy need a large array of complementary services such as: financial services, a highly educated workforce, sources of knowledge (universities and research centers), logistic services, etc. There is an avalanche of recent studies that confirm the above premises.

This space-time growth process is particularly evident in Germany where cities (and the surrounding districts) compete actively with one another for new business by using a portfolio of policy measures to create a business-friendly environment (Panebianco 2005). The special focus on Germany is – apart from its socio-political dynamics - warranted by its pro-active policy making at both regional and city (district) level (Panebianco et al. 2005). Our study aims to identify the central role of major employment centers in Germany.

We will consider German cities (as major cities in a labour market district) as central nodes in a socio-economic network. The agglomeration externalities then work towards the formation of hubs, which may be called ‘preferential nodes’. New entrepreneurs benefit by locating near hubs (attraction nodes) because these are the places where new opportunities emerge. In addition, from the hubs outwards, innovation may quickly reach dispersed and remote cities. This connectivity feature of city networks can be interpreted as a ‘scale-free network’ (Gorman 2005; and see Section 3).

The process of agglomeration through the economics of density is further reinforced by the fact that many German firms, especially small and medium enterprises (as a result of new entrepreneurial activity) arose out of regional networks where local banks, state and city governments played an important role (Kogut and Walker 2001).

We may therefore, argue that, if cities and districts act as hubs for entrepreneurial activity (because of the range of opportunities they offer), they should attract considerable labour inputs from outside, and possibly from quite far away. Against this backdrop, the aim of this paper is to explore the presence of hubs in the German system of regional market centers, by investigating the connectivity structure of the network itself.

The structure of the present paper is as follows. Section 2 introduces the characteristics of network connectivity models. Next, Section 3 presents the results of the empirical analysis, while Section 4 concludes the paper with interpretations and reflections on future research.

2. Network Models and Connectivity: A Brief Review

In recent years, great interest has arisen in the interdisciplinary study of complex networks, and in particular in the relevance of the interconnectivity structures. The strength of (dynamic) interactions and the form of the connectivity systems seem to be critical to identify the network properties and related complex dynamics (Reggiani and Nijkamp 2004).

In this framework, the concept of Scale-Free (SF) networks – originally introduced by (Barabási and Albert 1999) – has gained a great deal of attention, essentially for its interesting characteristic of exhibiting power-law distributions in the connectivity structures of a network. SF models have been embraced as generic, yet universal models of network topologies, and thus been suggested as representative models of complex systems, ranging from the social sciences to molecular biology, or to the internet (Alderson et al. 2004). Clear real-world examples of SF models have also been proven to exist in spatial-economic systems, such as in the telecommunication, transport and peer-to-peer networks (Gorman 2005, Gorman et al. 2005, Schintler et al. 2005).

The most prominent feature of SF networks – beyond exhibiting a power-law rank-size distribution – is the presence of highly connected nodes (hubs), as outlined by (Barabási and Oltvai (2004, p.104): "*Networks that are characterized by a power-law degree distribution are highly non-uniform, most of the nodes have only a few links. A few nodes with a very large number of links, which are often called hubs, hold these nodes together. Networks with a power degree distribution are called scale-free, a name that is rooted in statistical physics literature. It indicates the absence of a typical node in the network (one that could be used to characterize the rest of the nodes). This is in strong contrast to random networks, for which the degree of all nodes is in the vicinity of the average degree, which could be considered as typical.*"

This last sentence by Barabási and Oltvai summarizes the critical difference between SF and another common network model, i.e. the Random Network (RN). RN model – originally introduced by Erdos and Rényi (1959) – consists of nodes with randomly placed connections. In such a network, a plot of the distribution of node linkages follows a Poisson distribution (bell-shaped curve), which shows that most nodes have approximately the same number of links (i.e. the network average degree <*k*>). The tail of the degree distribution decreases exponentially, which shows that nodes that significantly deviate from the average <*k*> are very rare. Currently, there is a great deal of discussion on the precise definitions, rigorous proof of properties, and ubiquity of SF networks (Alderson et al. 2004). We will not engage here in such a debate, but rather start from a heuristic perspective by addressing a practical approach, i.e. the exploration of the type of connectivity structure of the commuting network in Germany. From this perspective, we will examine the possibility of a power-law degree distribution vs. an exponential-law distribution.

The exponential distribution is a rather simple function which has often emerged in the spatial economics literature. In the context of network analysis, the exponential distribution may be considered to belong to the class of RN models, since it suggests – with respect to the power-law distribution – that a multiplicity of nodes with a few links does not exist: in other words, the network is rather homogeneous and does not present a 'hub' structure. This latter structure might then show a spatially-equilibrated pattern with an exponential deterrence function. In other words, in the context of commuting it means an underrepresentation of commuters on long distances, and hence we might conjecture 'slow diffusion dynamics' in the peripheries for entrepreneurial start-ups. On the other hand, the power-law might show the existence of hubs, i.e. the relative irrelevance of distance, and thus 'fast diffusion dynamics', in the entrepreneurship process.

On the basis of the previous considerations, in the next section we will present the results of our empirical analysis devoted to the exploration the connectivity structure of the commuting network in Germany.

3. Empirical Analysis

In this exploratory analysis we investigate whether the real (commuting) network formed by the nodal structure of German labour market districts can be identified as an SF model. The nodes of the network under analysis are the 439 economic districts in Germany (Figure 1). The links between the nodes are the commuting flows between any two districts – the data refer to the year 2002.[1] The number of commuters on a given link will represent the strength of the link. Hence, a link of strength 1 connecting two districts implies that only one commuter travelled between the two districts during the year 2002.

[1] There is substantial commuting within districts. However, because the distribution of the distance travelled by commuters is not available, we have limited the analysis to commuting flows between districts.

FIGURE 1 ABOUT HERE

Commuting flows are an interesting variable for network analysis, because they can be regarded as a synthetic measure of the overall level of economic exchange between regions. The data under analysis show a spatial interaction model structure; i.e. a negative relationship between the number of commuters and the distance travelled (see also Gorman et al. 2005). In particular, on short links we find many commuters (the strength of the link is highest on short links) while on long-distance links there are few commuters (long links are usually weak, because few commuters are prepared to travel very long distances). On the longest distance, only one or only a few commuters are found.

It may be argued that a very low level of commuting flows, i.e. weak links, is not interesting from an economic point of view. However, since they are important for the test of SF theory, we decided to retain all links, even the weakest ones, i.e. those characterized by the strength of just one commuter. In fact, individuals may travel very long distances during a part of the year, while relocating their household. Nonetheless, we will check the robustness of our results vis-à-vis the exclusion of weak links.

We have ranked the German labour market districts by the number of incoming links (commuting flows; each incoming commuter flow represents a link with another district, i.e. the district where the flow of commuters originates). The highest rank is assigned to the district (positioned[2] as 1st) with the highest number of links.

The question whether a real-world network is a SF network boils down to the assessment of whether the relationship between the logarithm of the number of links and the rank of the district follows a power-law distribution rather than an exponential one (see Figure 2).

FIGURE 2 ABOUT HERE

To this end, we have run the following two regression analyses:

$$\ln(N_i) = \alpha + \beta R_i + \varepsilon_i \tag{1}$$

and

$$\ln(N_i) = \delta + \gamma \ln(R_i) + u_i\,, \tag{2}$$

where N_i is the of number of commuting flows with destination district i, R_i is the corresponding district rank, α, β, δ and γ are parameters to be estimated, and ε_i and u_i are two i.i.d. normally-distributed error components. Equation (1) derives from an exponential

[2] [2] Ties are broken randomly.

relationship between N and R ($N_i = e^{\alpha} e^{\beta R_i} e^{\varepsilon_i}$), while equation (2) implies that the relationship between R and N can be described by a power-law ($N_i = e^{\delta} R_i^{\gamma} e^{v_i}$).

The results of the estimation of these two models are presented in the upper panel of Table 1.

TABLE 1 ABOUT HERE

The comparison of the R^2s obtained from the two models shows that the exponential model performs better than the power-law model (in the sense that the exponential specification fits the data better than the power-law specification). However, this conclusion could be misleading. Because the power-law specification does not nest the exponential specification as a special case (and *vice versa*), the analysis of the R^2 criterion alone is not sufficient to warrant the rejection the power-law specification in favour of the exponential specification (and *vice versa*).[3]

The choice between non-nested competing models can be further highlighted by two statistical tests: the J test and an encompassing test (Davidson and MacKinnon 2004). These tests are presented in the lower panel in Table 1. The tests themselves are described in the statistical Annex A of the present paper.

The J test shows that neither model specification satisfactorily fits the data. In fact, both model specifications are rejected. Thus, although the exponential model fits the data better than the power-law model, the exponential model is still mis-specified. This inference is also supported by the results of the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test verifies (and in this case rejects) the equality between distributions; here we test whether the distribution of ln(N) can be considered equal to the distribution of the predicted values obtained from the exponential model ($\ln(N)^{ex}$), and to the distribution of the predicted values obtained from the power-law specification ($\ln(N)^{pl}$). Despite the higher R^2, it is quite difficult to argue that the relationship between the number of links and the district rank is captured by the exponential specification (as opposed to the power-law specification).

One could argue that weak links (links carrying just a few commuters over the year) may actually not be important from an economic point of view (because the impact of measurement errors could be severe). To this end, we have checked the robustness of our results against measurement errors in the flows of commuters by excluding the weakest links.

[3] The comparison between the R^2s can clarify the selection of competing nested models, but it cannot be used to make an inference about the functional form of the underlying true model (unless the R^2 is equal to 1). In fact, the R^2 is an indication of the significance of the parameters and the overall goodness of fit. To investigate the level of agreement between the functional form chosen and the data, a specification test, such as the Ramsey's RESET test, ought to be used.

In other words, we have re-run the analysis considering only those links carrying more than 10, 100, 1000, 5000, 10000, and 15000 commuters per year, respectively.[4] The results are presented in Table 2.

TABLE 2 ABOUT HERE

The first important result is that as, the number of links included decreases (and the minimum strength of the link increases), the performance of the power-law model – at least in terms of R^2 – appears to improve, in particular when only the links with more than 10000 commuters per year are included in the analysis. However, despite the increases in the R^2, the power-law model is not able to describe the data properly. The only exception to this remark arises when the analysis considers links with more than 5000 commuters per year; in this case, the power-law model appears to be superior to the exponential specification, and only the Kolmogorov-Smirnov test does not acknowledge this superiority.

The superiority of the power-law model with respect to the exponential specification is not robust, however: as the number of commuters per year increases, the superiority of the power-law model vanishes and both model specifications are rejected.

All in all, our results suggest that both the exponential and the power-law model specifications are not able to capture all salient features in the data.

Next, an inspection of the distribution of the log number of links against the district rank shown in Figure 2 suggests that both model specifications could be improved by the addition of a quadratic term.[5] Consequently, we have re-run the whole analysis using a quadratic specification[6]:

$$\ln(N_i) = \alpha + \beta_1 R_i + \beta_2 R^2{}_i + \varepsilon_i \qquad (3)$$

and

$$\ln(N_i) = \delta + \gamma_1 \ln(R_i) + \gamma_2 [\ln(R_i)]^2 + v_i \,. \qquad (4)$$

The results are presented in Table 3.

TABLE 3 ABOUT HERE

[4] As the required minimum strength of the link increases, a few districts (nodes) drop out from the sample.

[5] A regression line through the distribution would overestimate the log number of links for low and high values of a city rank, and it would underestimate the log number of links for intermediate values of a city rank.

[6] The choice for the quadratic specification has been guided by the estimation of a Box-Cox transformation $\frac{(district\ rank)^{\lambda} - 1}{\lambda}$. If λ =0, then the logarithmic transformation of the regressor district rank should be used. In our case, λ =1.87, which is statistically different from 0, thus implying that the logarithmic transformation is to be rejected. A value of λ close to 2 suggests a quadratic specification, however.

The quadratic specification significantly improves the fit of the model as shown by the comparison of the R^2s in Tables 1 and 3. This procedure is correct, because both quadratic models in Table 3 nest the linear models in Table 1 as special cases (when the coefficient of the quadratic term is zero).[7]

The comparison of the R^2s in the upper panel in Table 3 suggests that the exponential specification fits the data better than the power-law specification. This inference is supported by the additional tests presented. The (quadratic) exponential specification clears two out of three specification tests (the J test and the Kolmogorov-Smirnov test). The exponential specification fails to clear the encompassing test, but so does the power-law specification (the test rejects both the exponential and the power-low specifications). The power-law specification is soundly rejected. On the contrary, the F statistic corresponding to the exponential specification is low, significantly different from zero but low.

Table 4 shows that this result is again not very robust to the exclusion of the weakest links. As the required minimum strength of the link increases, the exponential specification ceases to be the preferred model specification.

TABLE 4 ABOUT HERE

All in all, neither the exponential nor the power-law specification can be regarded as superior. Moreover, in the absence of strong priors against the inclusion of all links (regardless of the strength of the link) and using a quadratic functional form, the (quadratic) exponential specification should be preferred to a (quadratic) power-law specification.

The poor performance of both model specifications may be ascribed to the inability of the models to capture the sharp drop in the log number of links when the district rank is above 403. Consequently, we reconsidered the performance of all four model specifications (exponential and power-law, linear and quadratic) when districts ranked above 403 are excluded from the analysis. The results of this final experiment are shown in Table 5.

TABLE 5 ABOUT HERE

---

[7] It is well known that the significance of the quadratic term in the log-log model shown in equation (4) will be found significant even if Zipf's law were to hold. By the same token, we are also aware that the use of OLS to estimate the parameters characterizing the power-law relationship between node rank and the corresponding number of links (Zipf's law) introduces an upward bias in the estimated value of γ in equation (2). We show the results anyway, by way of illustration of the working of the encompassing tests. On the one hand, the assessment of the validity of Zipf's law is a matter of estimation rather than testing (almost any specification, also the true one, will be rejected if the data points are sufficiently numerous); on the other hand, the decision between competing specifications must be guided by tests.

Although the observations in the right tails have not been included in the analysis, the linear exponential specification is by no means superior to the power-law specification. In addition, the comparison of the R^2s shows that both quadratic specifications perform significantly better than the corresponding linear specifications. Furthermore, among the quadratic specifications, the specification tests suggest that the (quadratic) exponential specification ($N_i = e^{\alpha} e^{\beta_1 R_i} e^{\beta_2 R_i^2} e^{\varepsilon_i}$) appears to be superior to the (quadratic power-law) specification ($N_i = e^{\delta} R_i^{\gamma_1} e^{\gamma_2 (\ln R_i)^2} e^{v_i}$).

On the whole, our results suggest that commuting flows form a rather homogeneous and spatially-equilibrated network. In other words, the network does not seem to be dominated by a few hubs (districts) with many links. This network feature could be due to a rather balanced regional development.

Finally, a *caveat* is in order. There are many reasons why the commuting network may fail to show up as a SF. First, the links concerned might, in fact, be poor representations of network connectivity. If commuting flows are not reliable indicators of overall economic flows, then we may fail to detect important hubs, even though the German spatial-economic system might be dominated by hubs (which we were not able to detect). In this context, freight/trade flows, financial flows, and ICT flows might be more useful indicators to detect hubs in spatial entrepreneurial activities.

Second, the concept of hubs may refer to firms rather than to districts. In this case, start-ups may benefit from setting up business relationships in the area concerned with well-established key firms, that are very well connected with the rest of the economy (i.e. a nested hierarchical hub structure).

4. Conclusions

In the present analysis we have addressed the question whether a real spatial network – such as the one constituted by German labour market districts as nodes and commuting flows as links – can be identified as SF (scale-free). SF networks are dominated by a few important nodes (hubs) that could function as incubators for entrepreneurial activity. In addition, given the property that SF networks are ultra-small (Reuven and Shiomo 2003), innovation diffusion by entrepreneurship may reach remote nodes very rapidly, while hubs may also be able to attract workers from a considerable distance. In a SF network, the relationship between the logarithm of the number of links and the rank of the districts (in terms of the number of links) is best described by a power-law specification. *Vice versa*, in a homogeneous spatial structure where flows decay with distance, the relationship between the logarithm of the number of links vs. the rank of the districts (in terms of the number of links) is better identified by an exponential distribution (where the tail approaches to zero).

Given the economic relevance of flows related to long distances, we analysed the whole network, also including links with very few commuters. Our results indicate that, because the power-law specification is hardly ever superior to the exponential distribution, the German commuter network cannot be considered a SF network. The lack of particularly important hubs may be due to the relatively homogeneous distribution of highly industrialized districts.

It goes without saying that the selection of the preferred specification depends on the network characteristics, which are in turn determined by the criteria used to select the links forming the network. Finally, our analysis also makes clear that the R^2 alone is not a sufficient criterion to discriminate between competing model specifications, in particular when the competing hypotheses are non-nested.

The analysis could be extended in many directions. For example, the analysis of the outgoing commuting flows could help unravel the characteristics of (technological) innovation diffusion processes. This interesting application is not pursued here and is left for future research.

Future research efforts may also be directed – on the one hand – to the exploration of changes in the connectivity structure of commuting flows over time, by analyzing different time periods; on the other hand, they may also address the investigation of German networks that are more directly related to entrepreneurship features, such as ICT, telecommunication and/or freight/trade networks.

Acknowledgment

The authors wish to thank Günter Haag (STASA, Stuttgart, Germany) for kindly providing the data set used in the present study.

References

Alderson, L. L., Reisko, T., Doyle, J. C. and Willinger, W. (2004) "Towards a theory of scale-free graphs: definition, properties, and implications", California Institute of Technology, Engineering & Applied Sciences Division, Technical Report CIT-CDS-04-006.

Audretsch, D. B. (1995) *Innovation and industry evolution*, Cambridge, MIT Press.

Audretsch, D. B. (2004) "Sustaining innovation and growth: public policy support for entrepreneurship" *Industry and Innovation*, 11, 167-191.

Audretsch, D. B. and Thurik, R. (2001) "Linking entrepreneurship to growth", OECD STI Working Paper, No 2001/2.

Audretsch, D. B. and Fritsch, M. (2002) "Growth regimes over time and space" *Regional Studies*, 36, 113-124.

Audretsch, D. B. and Fritsch, M. (2003) "Linking entrepreneurship to growth: the case of West Germany" *Industry and Innovation*, 10, 65-73.

Barabási, A.-L. L. and Albert, R. (1999) "Emergence of Scaling in Random Networks" *Science*, 286, 509-512.
Barabási, A.-L. L. and Oltvai, Z. N. (2004) "Networks Biology: Understanding the Cell's Functional Organization" *Nature Reviews - Genetics*, 5, 101-113.
Carree, M. A. and Thurik, R. (2002) "The impact of entrepreneurship on economic growth", Erasmus University Rotterdam, Mimeo.
Caves, R. E. (1998) "Industrial Organization and new findings on the turnover and mobility of firms" *Journal of Economic Literature*, 36, 1947-1982.
Davidson, R. and MacKinnon, J. G. (2004) *Econometric theory and methods*, New York, Oxford University Press.
Davidsson, P. and Henrekson, M. (2002) "Determinance of the prevalence of start-ups and high-growth firms" *Small Business Economics*, 19, forthcoming.
Erdos, P. and Rényi, A. (1959) "On random graphs" *Publ. Math*, 9, 290-297.
Glaeser, E. L. (1998) "Are cities dying?" *Journal of Economic Perspectives*, 12, 139-160.
Gorman, S. (2005) *Networks, security and complexity*, Cheltenham, UK, Edward Elgar.
Gorman, S., Patuelli, R., Reggiani, A., Nijkamp, P., Kulkarni, R. and Haag, G. (2005) "An application of complex network theory to German commuting patterns", Paper presented at the NSF Workshop, Penn State University, May 9-11
Greene, W. H. (2003) *Econometric analysis*, Upper Saddle River, Prentice Hall.
Jorgenson, D. W. and Stiroh, K. J. (2000) "Raising the speed limit: US economic growth in the information age" *Brookings Papers on Economic Activity*, 1, 125 - 211.
Jorgenson, D. W., Ho, M. S. and Stiroh, K. J. (2003) "Lessons from the US growth resurgence" *Journal of Policy Modeling*, 25, 453 - 470.
Kogut, B. and Walker, G. (2001) "The small world of Germany and the durability of national networks" *American Sociological Review*, 66, 317-335.
Mizon, G. E. and Richard, J.-F. (1986) "The encompassing principle and its application to testing non-nested hypotheses" *Econometrica*, 54, 657-678.
OECD (2003) *The sources of economic growth in OECD countries*, Paris, OECD.
Panebianco, S. (2005) "Are entrepreneurial cities more successful? Empirical evidence from 50 German cities", Paper Presented at COST A26, Dortmund, 16-18 June, 2005.
Panebianco, S., Fasselt, J., Doenitz, U. and Zimmer-Hegmann, R. (2005) "Evaluating regional governance: methodological concerns and practical experiences", Paper Presented at ERSA 2005, Amsterdam, 27-31 August 2005.
Reggiani, A. and Nijkamp, P. (2004) "The fall of the iron curtain and the evolution of regional labour markets: a self-organised criticality perspective" in O. Yasuhide and S. E. Chang *Modeling spatial and economic impacts of disasters*, Berlin, Springer-Verlag.
Reuven, C. and Shiomo, H. (2003) "Scale-free networks are ultrasmall" *Physical Review Letters*, 90, 058701-058701/058704.
Reynolds, P. D. (1999) "Creative destruction: source or symptom of economic growth?" in Z. J. Acs, B. Carlsson and C. Karlsson *Entrepreneurship, small and medium-sized enterprises and the Macroeconomy*, Cambridge, Cambridge University Press.
Schintler, L., Gorman, S. P., Reggiani, A., Patuelli, R. and Nijkamp, P. (2005) "Small world phenomena in communication networks: a cross-Atlantic comparison" in A. Reggiani and L. Schintler *Methods and models in transport and telecommunications: cross Atlantic perspectives*, Berlin, Springer Verlag.
van Stel, A. and Diephuis, B. (2004) "Business dynamics and employment growth: a cross-country analysis", Discussion Paper on Entrepreneurship, Growth and Public Policy, No 3204.
Sutton, J. (1997) "Gibrat's legacy" *Journal of Economic Literature*, 35, 40-59.

ANNEX A: Testing Non-nested Hypotheses

The comparison between R^2s can guide the researcher through the choice between competing nested models but it cannot be used to assess the validity of competing functional form. An example can clarify this point.

Assume the data are generated by a quadratic data-generating process ($y_i = \theta x_i^2 + \varsigma_i$). Assume further that we may choose to compare the following two models: model A: $y_i = \alpha + \varepsilon_i$, and model B: $y_i = \alpha + \beta x_i + \varepsilon_i$. Clearly, $R^2_B > R^2_A$. However, it would be incorrect to conclude that the data-generating process is linear. Similarly, the R^2 criterion cannot be used to choose between non-nested hypotheses.

In traditional testing, the H_0 can be expressed as a particular case of the more general H_1 (the alternative hypothesis); in this sense the H_0 is nested in the H_1. When hypotheses are non-nested the H_0 and H_1 are on equal footing. The small sample and the asymptotic proprieties of the tests described here have received a thorough treatment in (Davidson and MacKinnon 2004, Greene 2003, Mizon and Richard 1986).

There are two possible approaches to test non-nested hypotheses: a) the comprehensive approach (the J-test), and b) the encompassing approach. We begin with the former.

a) The J-Test

Suppose we have to decide between two rival models:

$$H_0: \qquad y_i = \beta' X_i + v_i \tag{A1}$$

and

$$H_1: \qquad y_i = \gamma' Z + u_i , \tag{A2}$$

where β (p x 1) and γ (q x 1) are vectors of parameters to be estimated, X (p x n) and Z (q x n) matrices of exogenous variables, and u and v are two error terms that follow the usual assumptions.

We notice that the variables included in Z and X should be different and it should not be the case that X (Z) could be obtained as a special case of Z (X), i.e. setting some parameters to zero.

The J-test prescribes the following steps:

1. estimate equation (A1) and obtain the predicted values $\hat{y}^X = \hat{\beta}' X$;
2. estimate equation (A2) and obtain the predicted values $\hat{y}^Z = \hat{\gamma}' Z$;
3. estimate $y_i = \beta' X_i + \delta \hat{y}_i^z + v_i$; if δ=0, the model in equation (A1) cannot be rejected; if δ≠ 0, then the model in equation (A1) can be rejected (the Z variables

still incorporate some features that help improve on the performance of model (A1));

4. estimate $y_i = \beta' X_i + \pi \hat{y}_i^X + u_i$; if π=0, the model in equation (A2) cannot be rejected; if π≠ 0, then the model in equation (A2) can be rejected (the X variables incorporate some features that help improve on the performance of model (A2)).

In other words, the J-test boils down to a t-test on the additional variable represented by the predicted values of the rival model. The estimates of δ and π and their associated standard errors are reported in Tables 1 - 5.

b) The Encompassing Test

Suppose that, again, we have to decide between the two aforementioned models. The encompassing test prescribes that the variables be divided into three groups: the matrix W that contains the common regressors (i.e. regressors that belong to both Z and X); the matrix Z^n that includes all exogenous variables present in the matrix Z but not in the matrix X; finally, the matrix X^n that includes all the regressors present in the matrix X but not in the matrix Z.

Then the rival models can be combined into a super-model (a model that encompasses both sub-models as special cases):

$$y_i = \alpha' W_i + \beta' X_i^n + \gamma' Z_i^n + u_i, \qquad \text{(A3)}$$

where α, β and γ are vectors of parameters to be estimated, and u is a random disturbance that follows the usual assumptions.

The encompassing test prescribes that H_1 is rejected if γ=0, and H_0 is rejected if β=0. These restrictions can be tested by means of F-tests. Sometimes Z^n or X^n are vectors (nx1), i.e. they just include one variable. In this case the F-test can be approximated by the squared value of the t-test obtained during the estimation of equation (A3). The values of the F-test (or the t-test when needed) and their relative p-values (or standard errors in the case of a t-test) are reported in Tables 1 - 5.

Both the J-test and the encompassing test have four possible outcomes:

1. Reject H_0 and do not reject H_1;
2. Do not reject Ho and reject H_1;
3. Reject H_0 and reject H_1;
4. Do not reject H_0 and do not reject H_1.

In the first two cases, one of the two models ends up as the preferred model. In the third case, neither model is correctly specified. In the fourth case, both models are correctly specified.

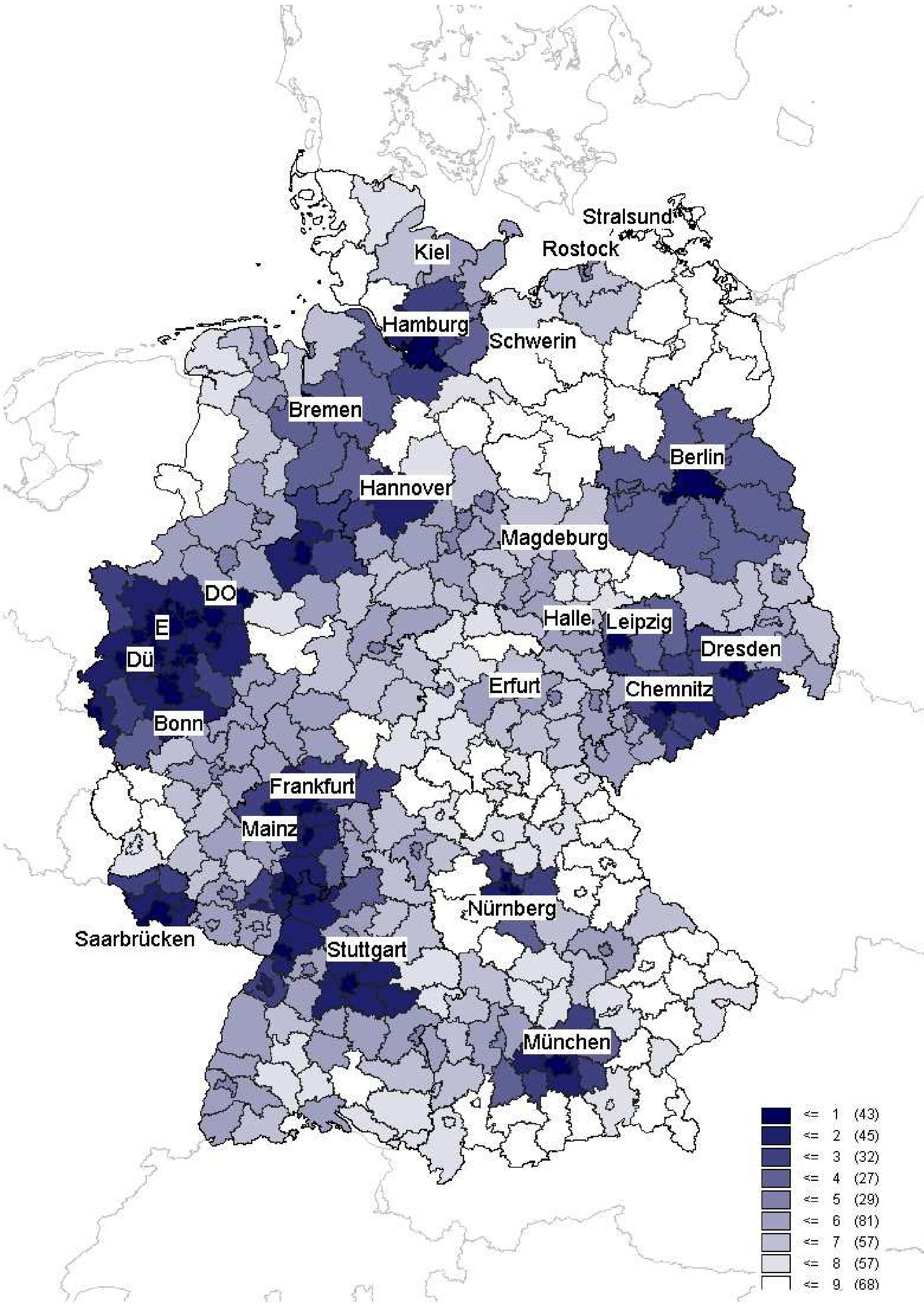

Figure 1: The regional distribution of the 439 districts in Germany

(Legend: 1. Central cities in regions with urban agglomerations; 2. Highly urbanized districts in regions with urban agglomerations; 3. Urbanized districts in regions with urban agglomerations; 4. Rural districts in regions with urban agglomerations; 5. Central cities in regions with tendencies towards agglomeration; 6. Highly urbanized districts in regions with tendencies towards agglomeration; 7. Rural districts in regions with tendencies towards agglomeration; 8. Urbanized districts in regions with rural features; 9. Rural districts in regions with rural features.)

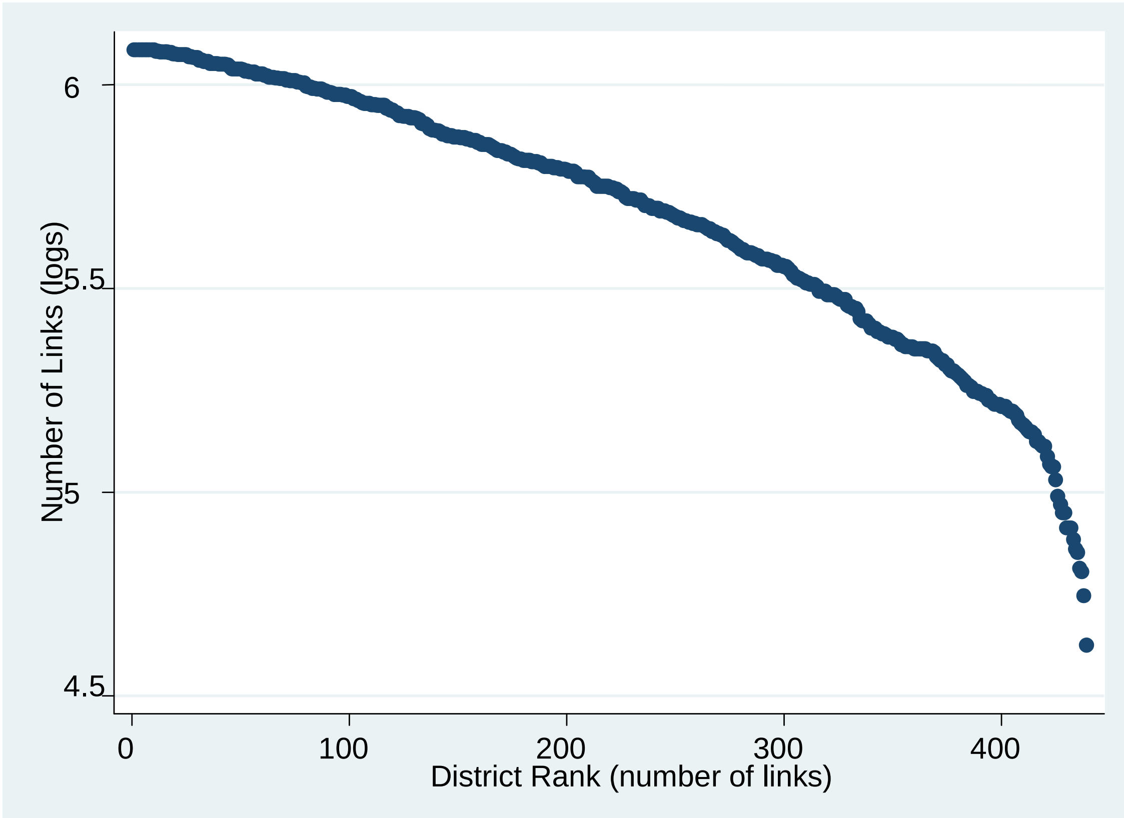

Figure 2: The relationship between the logarithm of the number of links and the district rank – commuting network in Germany (in 2002)

Table 1: Regression analyses, linear model specifications in equation (1) and equation (2). The whole commuting network in Germany (standard errors in brackets and p-values in parenthesis. The test is significant at the 5% confidence level when the ratio of the coefficient to the standard error is larger than 1.96 and when the p-value is smaller than 0.05).

	Exponential	Power-Law
Constant	6.211	6.976
	[0.001]	[0.050]
Ln District Rank		-0.254
		[0.010]
District Rank	-0.002	
	[0.000]	
R^2	0.941	0.616
number of cases	439	439

Specification Tests

J-Test

H_0	H_1	T-Test Statistic	Result
Power-Law Specification	Exponential Specification	1.268 [0.020]	Reject H_0
Exponential Specification	Power-Law Specification	-0.377 [0.025]	Reject H_0

Encompassing Test

H_0	H_1	T-Test Statistic	Result
The Data Generating Process Follows Power-Law	The Data-Generating Process Follows Exponential Specification	-0.003 [0.000]	Reject H_0
The Data-Generating Process Follows Exponential Specification	The Data-Generating Process Follows Power-Law	0.096 [0.006]	Reject H_0

Kolmogorov - Smirnov Test

H_0	H_1	KS-Test Statistic	Result
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{pl}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{pl}$	0.239 (0.000)	Reject H_0
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{ex}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{ex}$	0.119 (0.004)	Reject H_0

Table 2: Robustness of the two linear models to different network structures (standard errors in brackets and p-values in parenthesis. The test is significant at the 5% confidence level when the ratio of the coefficient to the standard error is larger than 1.96 and when the p-value is smaller than 0.05).

Minimum strength of the Link:		10		100		1000	
Regression		R^2		R^2		R^2	
Exponential		0.816		0.969		0.894	
Power-Law		0.969		0.815		0.851	
Number of Cases		439		439		439	
Specification Tests							
J-Test							
H_0	H_1	T-Test Statistic	Result	T-Test Statistic	Result	T-Test Statistic	Result
Power-Law Specification	Exponential Specification	0.186 [0.017]	Reject H_0	0.186 [0.017]	Reject H_0	0.625 [0.028]	Reject H_0
Exponential Specification	Power-Law Specification	0.851 [0.016]	Reject H_0	0.851 [0.016]	Reject H_0	0.437 [0.029]	Reject H_0
Encompassing Test							
H_0	H_1	T-Test Statistic	Result	T-Test Statistic	Result	T-Test Statistic	Result
The Data-Generating Process Follows Power-Law	The Data-Generating Process Follows Exponential Specification	-0.529 [0.010]	Reject H_0	-0.529 [0.009]	Reject H_0	-0.255 [0.017]	Reject H_0
The Data-Generating Process Follows Exponential Specification	The Data-Generating Process Follows Power-Law	-0.0008 [0.0001]	Reject H_0	-0.0008 [0.0001]	Reject H_0	-0.003 [0.000]	Reject H_0
Kolmogorov - Smirnov Test							
H_0	H_1	KS-Test Statistic	Result	KS-Test Statistic	Result	KS-Test Statistic	Result
Cumulative Distribution of ln(N) = Cumulative Distrbution of $\ln(N)^{pl}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{pl}$	0.118 (0.000)	Reject H_0	0.118 (0.000)	Reject H_0	0.223 (0.000)	Reject H_0
H_0	H_1						
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{ex}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{ex}$	0.143 (0.000)	Reject H_0	0.143 (0.000)	Reject H_0	0.155 (0.004)	Reject H_0

Table 2: Continued

Minimum Strength of the Link:		5000		10000		15000	
Regression		R^2		R^2		R^2	
Exponential		0.695		0.432		0.273	
Power-Law		0.909		0.814		0.694	
Number of Cases		439		418		377	
Specification Tests							
J-Test							
H_0	H_1	T-Test Statistic	Result	T-Test Statistic	Result	T-Test Statistic	Result
Power-Law Specification	Exponential Specification	-0.020 [0.036]	Do Not Reject H_0	-0.906 [0.051]	Reject H_0	-1.779 [0.069]	Reject H_0
Exponential Specification	Power-Law Specification	1.015 [0.032]	Reject H_0	1.580 [0.037]	Reject H_0	1.982 [0.043]	Reject H_0
Encompassing Test							
H_0	H_1	T-Test Statistic	Result	T-Test Statistic	Result	T-Test Statistic	Result
The Data-Generating Process Follows Power-Law	The Data-Generating Process Follows Exponential Specification	0.0001 [0.0001]	Do Not Reject H_0	0.002 [0.000]	Reject H_0	-0.003 [0.000]	Reject H_0
The Data-Generating Process Follows Exponential Specification	The Data-Generating Process Follows Power-Law	-0.563 [0.018]	Reject H_0	-0.567 [0.013]	Reject H_0	-0.521 [0.011]	Reject H_0
Kolmogorov - Smirnov Test							
H_0	H_1	KS-Test Statistic	Result	KS-Test Statistic	Result	KS-Test Statistic	Result
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{pl}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{pl}$	0.369 (0.000)	Reject H_0	0.404 (0.000)	Reject H_0	0.453 (0.000)	Reject H_0
H_0	H_1						
Cumulative Distribution of ln(N) = Cumulative Distrbution of $\ln(N)^{ex}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{ex}$	0.362 (0.000)	Reject H_0	0.500 (0.000)	Reject H_0	0.567 (0.000)	Reject H_0

Table 3: Regression analyses, quadratic model specifications in equation (3) and equation (4). The whole commuting network in Germany (standard errors in brackets and p-values in parenthesis. The test is significant at the 5% confidence level when the ratio of the coefficient to the standard error is larger than 1.96 and when the p-value is smaller than 0.05).

	Exponential	Power-Law
Constant	6.062 [0.006]	5.368 [0.066]
Ln District Rank		0.584 [0.031]
(Ln District Rank)^2		-0.099 [0.004]
District Rank	-0.0004 [0.0001]	
District Rank^2	-0.00001 [0.0000001]	
R^2	0.985	0.858
Number of Cases	439	439

Specification Tests

J-Test

H_0	H_1	T-Test Statistic	Result
Power-Law Specification	Exponential Specification	1.006 [0.016]	Reject H_0
Exponential Specification	Power-Law Specification	-0.010 [0.032]	Do Not Reject H_0

Encompassing Test

H_0	H_1	F-Test Statistic	Result
The Data-Generating Process Follows Power-Law	The Data-Generating Process Follows Exponential Specification	2308.499 (0.000)	Reject H_0
The Data-Generating Process Follows Exponential Specification	The Data-Generating Process Follows Power-Law	48.741 (0.000)	Reject H_0

Kolmogorov - Smirnov Test

H_0	H_1	KS-Test Statistic	Result
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{pl}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{pl}$	0.132 (0.001)	Reject H_0
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{ex}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{ex}$	0.068 (0.257)	Do Not Reject H_0

Table 4: Robustness of the two quadratic models to different network structures (standard errors in brackets and p-values in parenthesis. The test is significant at the 5% confidence level when the ratio of the coefficient to the standard error is larger than 1.96 and when the p-value is smaller than 0.05).

Minimum Strength of the Link:		10		100		1000	
Regression		R^2		R^2		R^2	
Exponential		0.989		0.816		0.896	
Power-Law		0.985		0.957		0.920	
Number of Cases		439		439		439	
Specification Tests							
J-Test							
H_0	H_1	T-Test Statistic	Result	T-Test Statistic	Result	T-Test Statistic	Result
Power-Law Specification	Exponential Specification	1.391	Reject H_0	-2.418	Reject H_0	2.345	Reject H_0
		[0.098]		[0.200]		[0.121]	
Exponential Specification	Power-Law Specification	-0.763	Reject H_0	1.851	Reject H_0	-7.542	Reject H_0
		[0.155]		[0.087]		[0.545]	
Encompassing Test							
H_0	H_1	F-Test Statistic	Result	F-Test Statistic	Result	F-Test Statistic	Result
The Data-Generating Process Follows Power-Law	The Data-Generating Process Follows Exponential Specification	199.88 (0.000)	Reject H_0	145.44 (0.000)	Reject H_0	377.380 (0.000)	Reject H_0
The Data-Generating Process Follows Exponential Specification	The Data-Generating Process Follows Power-Law	1308.56 (0.000)	Reject H_0	1674.66 (0.000)	Reject H_0	196.470 (0.000)	Reject H_0
Kolmogorov - Smirnov Test							
H_0	H_1	KS-Test Statistic	Result	KS-Test Statistic	Result	KS-Test Statistic	Result
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{pl}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{pl}$	0.062 (0.377)	Do Not Reject H_0	0.123 (0.003)	Reject H_0	0.203 (0.000)	Reject H_0
H_0	H_1						
Cumulative Distribution of ln(N) = Cumulative Distrbution of $\ln(N)^{ex}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{ex}$	0.052 (0.583)	Do Not Reject H_0	0.134 (0.001)	Reject H_0	0.180 (0.000)	Reject H_0

Table 4: Continued

Minimum Strength of the Link:		5000		10000		15000	
Regression		R^2		R^2		R^2	
Exponential		0.912		0.861		0.866	
Power-Law		0.928		0.903		0.847	
Number of Cases		439		418		377	
Specification Tests							
J-Test							
H_0	H_1	T-Test Statistic	Result	T-Test Statistic	Result	T-Test Statistic	Result
Power-Law Specification	Exponential Specification	1.208	Reject H_0	0.826	Reject H_0	0.480	Reject H_0
		[0.119]		[0.059]		[0.053]	
Exponential Specification	Power-Law Specification	-0.361	Reject H_0	0.245	Reject H_0	0.625	Reject H_0
		[0.166]		[0.067]		[0.051]	
Encompassing Test							
H_0	H_1	F-Test Statistic	Result	F-Test Statistic	Result	F-Test Statistic	Result
The Data-Generating Process Follows Power-Law	The Data-Generating Process Follows Exponential Specification	102.981	Reject H_0	197.903	Reject H_0	82.313	Reject H_0
		(0.000)		(0.000)		(0.000)	
The Data-Generating Process Follows Exponential Specification	The Data-Generating Process Follows Power-Law	710.874	Reject H_0	1043.513	Reject H_0	1049.271	Reject H_0
		(0.000)		(0.000)		(0.000)	
Kolmogorov - Smirnov Test							
H_0	H_1	KS-Test Statistic	Result	KS-Test Statistic	Result	KS-Test Statistic	Result
Cumulative Distribution of ln(N) = Cumulative Distribution of $\ln(N)^{pl}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{pl}$	0.353	Reject H_0	0.426	Reject H_0	0.528	Reject H_0
		(0.000)		(0.000)		(0.000)	
H_0	H_1						
Cumulative Distribution of ln(N) = Cumulative Distrbution of $\ln(N)^{ex}$	Cumulative Distribution of ln(N) ≠ Cumulative Distribution of $\ln(N)^{ex}$	0.355	Reject H_0	0.421	Reject H_0	0.496	Reject H_0
		(0.000)		(0.000)		(0.000)	

Table 5: Robustness of all four model specifications to different network structures – district rank < 404 (standard errors in brackets and p-values in parenthesis. The test is significant at the 5% confidence level when the ratio of the coefficient to the standard error is larger than 1.96 and when the p-value is smaller than 0.05).

Regressions	**Linear:**	Exponential	Power-Law	**Quadratic:** Exponential	Power-Law
Constant		6.176	6.811	6.086	5.547
		[0.004]	[0.039]	[0.001]	[0.046]
Ln District Rank			-0.214		0.459
			[0.008]		[0.022]
(Ln District Rank)^2					-0.081
					[0.003]
District Rank		-0.002		-0.001	
		[0.000]		[0.000]	
District Rank^2				-0.000003	
				[0.000000]	
R^2		0.974	0.661	0.999	0.902
Number of Cases		403	403	403	403
Specification Tests					
J-Test					
H_0	H_1	T-Test Statistic	Result	T-Test Statistic	Result
Power-Law Specification	Exponential Specification	1.217	Reject H_0	0.996	Reject H_0
		[0.012]		[0.005]	
Exponential Specification	Power-Law Specification	-0.299	Reject H_0	-0.013	Do Not
		[0.014]		[0.010]	Reject H_0
Encompassing Test					
H_0	H_1	T-Test Statistic	Result	F-Test Statistic	Result
The Data-Generating Process	The Data-Generating Process	-0.003	Reject H_0	21788.391	Reject H_0
Follows Power-Law	Follows Exponential	[0.000]		(0.000)	
The Data-Generating Process	The Data-Generating Process	0.064	Reject H_0	64.111	Reject H_0
Follows Exponential	Follows Power-Law	[0.003]		(0.000)	
Kolmogorov - Smirnov Test					
H_0	H_1	KS-Test Statistic	Result	KS-Test Statistic	Result
Cumulative Distribution of ln(N) =	Cumulative Distribution of ln(N) ≠	0.241	Reject H_0	0.141	Do Not
Cumulative Distribution of $\ln(N)^{pl}$	Cumulative Distribution of $\ln(N)^{pl}$	(0.000)		(0.001)	Reject H_0
Cumulative Distribution of ln(N) =	Cumulative Distribution of ln(N) ≠	0.104	Reject H_0	0.030	Reject H_0
Cumulative Distribution of $\ln(N)^{ex}$	Cumulative Distribution of $\ln(N)^{ex}$	(0.025)		(0.994)	